Assembly of Two Transgenes in an Artificial Chromatin Domain Gives Highly Coordinated Expression in Tobacco

Mlynárová, Ludmila; Loonen, Annelies; Mietkiewska, Elzbieta; Jansen, Ritsert C.; Nap, Jan-Peter

Published in:
Genetics

Assembly of Two Transgenes in an Artificial Chromatin Domain Gives Highly Coordinated Expression in Tobacco

Ludmila Mlynárová,*,† Annelies Loonen,* Elzbieta Mietkiewska,*,‡ Ritsert C. Jansen*,§ and Jan-Peter Nap*,[1]

**Plant Research International, Wageningen University and Research Center, NL-6700 AA Wageningen, The Netherlands, †Institute of Plant Genetics and Biotechnology, SK-950 07 Nitra, Slovak Republic, ‡Plant Breeding and Acclimatisation Institute, Młochów Research Center, PL-05 832 Rozalin, Poland and §Biometris, Wageningen University and Research Center, NL-6700 AA Wageningen, The Netherlands*

Manuscript received July 3, 2001
Accepted for publication December 4, 2001

ABSTRACT

The chromatin loop model predicts that genes within the same chromatin domain exhibit coordinated regulation. We here present the first direct experimental support for this model in plants. Two reporter genes, the *E. coli* β-glucuronidase gene and the firefly luciferase gene, driven by different promoters, were placed between copies of the chicken lysozyme A element, a member of the matrix-associated region (MAR) group of chromatin boundary elements, and introduced in tobacco (*Nicotiana tabacum*). In plants carrying A elements, quantitative enzyme activities and mRNA levels of both genes show high correlations compared to control plants. The A element thus creates an artificial chromatin domain that yields coordinated expression. Surprisingly, enzyme activities correlated poorly with their respective mRNA levels. We hypothesize that this indicates the occurrence of "error pipelines" in data generation: systematic errors of a given analytical method will point in the same direction and cancel out in correlation analysis, resulting in better correlations. In combining different methods of analysis, however, such errors do not cancel out and as a result relevant correlations can be masked. Such error pipelines will have to be taken into account when different types of (*e.g.*, whole-genome) data sets are combined in quantitative analyses.

A current model for the regulation of gene expression considers coordination in terms of functional units or modules. The concept of functional units of regulation supposes the presence of chromatin loop domains (Laemmli *et al.* 1992; Bode *et al.* 1996), delimited by sequences known as chromatin boundaries. Such boundaries create delimited chromatin loops and are thought to create topologically isolated units of gene regulation that shield and insulate genes located on the loop from *cis*-acting elements and other gene-repressing influences of the neighboring chromatin.

A whole-genome correlation analysis of yeast expression data indicated the existence of chromosomal domains of gene expression (Cohen *et al.* 2000), which further links structural domains with functional domains. Notably in mammalian systems, it is well established that multigene families, such as the β-globin locus, are arranged in well and linear-time-regulated units. In plants, it is suggested that regulation and development may be less linear (Scheres 2000) and could follow a more fuzzy logic. It is of particular interest, therefore, to know whether the concept of functional units of regulation can be applied to plants as well. Previously we have demonstrated a significant reduction in between-transformant variation and position-independent expression of the β-glucuronidase (GUS) reporter gene in transformed tobacco plants when the chicken lysozyme A element was placed at the borders of the T-DNA (Mlynárová *et al.* 1994, 1995, 1996). The reduced position effect of a single gene suggests the coincidence of structural loops with domains that reflect the functional organization of genome information. The chicken lysozyme A element is a member of the class of chromatin boundary elements known as matrix-associated regions (MARs). MAR elements are supposed to interact directly with the nuclear matrix, but the exact mechanism by which such elements achieve that influence is as yet unknown (Holmes-Davis and Comai 1998). Overall, a range of effects of MAR elements has been reported on gene expression levels and/or variation or stability in expression of transgenes in plants (Breyne *et al.* 1992; Allen *et al.* 1993, 1996, 2000; van der Geest *et al.* 1994; Han *et al.* 1997; Vain *et al.* 1999). Although the presence of MARs may protect against transcriptional silencing induced by *cis* interactions within repeated transgene arrays (Allen *et al.* 2000), their presence does not necessarily protect against the influence of strong silencing loci (Nap *et al.* 1997; Vaucheret *et al.* 1998). Also studies in cell lines demonstrate that not all MARs have the same mode of action, if any (Poljak *et al.* 1994; Dillon and Sabbattini 2000). The different effects are likely to arise from puta-

[1]*Corresponding author:* BU Genomics, Plant Research International, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands. E-mail: j.p.h.nap@plant.wag-ur.nl

tive differences between MARs in combination with different assay systems, transformation approaches, and recipient cells.

Until now, only single genes have been evaluated in functional MAR assays in plants. So far, no regulatory unit of gene regulation has been demonstrated in plants. This would require us to show that two individual genes linked on a single MAR-delimited chromatin loop become coordinately regulated. Such an approach is taken here. We investigate the creation of a unit of gene regulation by focusing on the coordination of the expression of two reporter genes assembled in a single T-DNA. The *Escherichia coli* GUS and the firefly luciferase (LUC) reporter genes, driven by two different promoters, were used. Quantitative analyses of gene expression demonstrate a significant correlation between the enzymatic activity levels as well as mRNA amounts of both reporter genes only in plants where the reporter genes are embedded in A-element DNA. This demonstrates the creation of an A-element-mediated unit of transcriptional regulation in plants and adds a new tool to the approaches available for achieving or studying coordinated expression of multiple genes in plants.

MATERIALS AND METHODS

Plant transformation and genetic analysis: Tobacco (*Nicotiana tabacum* cv Petit Havana SR1) was grown and transformed with *Agrobacterium tumefaciens* LBA4404 harboring the binary plasmid of interest as described previously (Mlynárová *et al.* 1994). During transformation, a selection pressure of 50 μg/ml kanamycin was applied. Rooting on hygromycin was assayed on root-inducing medium (MS medium supplemented with 0.05 μg/ml of IAA) with 20 μg/ml hygromycin (Duchefa). Plant populations were named after the vector they contain. Transgenic tobacco plants of interest were selfed. For each plant, ~40–50 offspring seeds were analyzed for segregation by germination without sterilization on quarter-strength MS medium supplemented with 50 μg/ml kanamycin or 20 μg/ml hygromycin as described previously (Conner *et al.* 1998).

Vectors for plant transformation: Standard procedures were used for DNA cloning and analysis (Sambrook *et al.* 1989). To facilitate cloning, the various DNA cassettes required for the binary vectors used in this study [MAR, GUS, LUC, hygromycin phosphotransferase (HPT)] were blunted, cloned in the standard cloning vectors pUC18 (New England BioLabs, Beverly, MA) or pSK+ (Stratagene, La Jolla, CA) and both orientations were identified. The chicken lysozyme A element was isolated as a *Bam*HI-*Xba*I fragment from plasmid pUC-B1-X1 (Phi-Van and Strätling 1988), blunted, and cloned into the *Sma*I site of pUC18. The GUS gene fused to the potato light-regulated *Lhca3.St.1* promoter and *nos* terminator was isolated as an *Xba*I-*Eco*RI fragment from plasmid pPPG (Nap *et al.* 1993a). This GUS cassette was blunted and cloned in the *Eco*RV site of pSK+. The gene for the cytosolic form of firefly luciferase was obtained from plasmid pSP-luc+ (Promega, Madison, WI) and supplied with the single 35S CaMV promoter with an N-terminal SV40 nuclear localization signal (NLS; van der Krol and Chua 1991) and the *nos* terminator. The presence of the NLS was shown to have no effect on the activity of the luciferase gene (van Leeuwen *et al.* 2000). The LUC cassette was isolated as a *Bam*HI-*Sst*I fragment, blunted, and cloned in the *Eco*RV site of pSK+. The hygromycin phosphotransferase gene driven by the *nos* promoter and carrying the polyadenylation signal pAg4 was obtained from plasmid pPCV720 (Koncz *et al.* 1994). This HPT cassette was isolated as a *Hin*dIII-*Asp*718 fragment, blunted, and cloned in the *Eco*RV site of pSK+. All plant transformation vectors were prepared in the binary vector pBinPLUS (van Engelen *et al.* 1995), which is a pBIN19-derivative with the *nos* promoter-driven kanamycin resistance selectable marker gene. Compared to pBIN19, pBinPLUS also has additional *Asc*I and *Pac*I cloning sites. The BinPLUS multiple cloning sites are also present in the pUC19-based plasmid pUCAP (van Engelen *et al.* 1995). The binary vectors prepared are named according to the T-DNA configuration they contain. To obtain pHGL, the HPT cassette as *Bam*HI-*Sal*I fragment was ligated to the GUS cassette as *Sal*I-*Eco*RI fragment and the LUC cassette as *Eco*RI-*Xba*I fragment in *Bam*HI-*Xba*I-digested pBinPLUS. The two MAR-element-containing plasmids pAHLGA and pAHGLA were made in two steps. For pAHLGA, the subclone pLGA was prepared in pUCAP by cloning the LUC cassette as *Hin*dIII-*Eco*RI fragment, the GUS cassette as *Eco*RI-*Kpn*I fragment, and the MAR cassette as *Kpn*I-*Xba*I fragment. The second subclone pAH was made by cloning the MAR cassette as *Kpn*I-*Bam*HI and the HPT cassette as *Bam*HI-*Sal*I fragment in pUCAP. Subsequently, the binary vector pAHLGA was obtained by ligation of the *Kpn*I-*Asc*I fragment from the pAH subclone and *Asc*I-*Xba*I fragment from the pLGA subclone in pBinPLUS. A similar strategy was used to obtain pAHGLA. Subclone pGLA was prepared in pUCAP by cloning the GUS cassette as a *Sal*I-*Eco*RI fragment, the LUC cassette as an *Eco*RI-*Kpn*I fragment, and the MAR cassette as a *Kpn*I-*Xba*I fragment. The pAHGLA binary vector was obtained by cloning the *Kpn*I-*Asc*I fragment from subclone pAH and the *Asc*I-*Xba*I fragment from subclone pGLA in pBinPLUS. All binary plasmids were conjugated to *A. tumefaciens* LBA4404 and verified by reintroduction into *E. coli* as previously described (Mlynárová *et al.* 1994). The structure of the T-DNA configurations after integration into the plant genome is shown in Figure 1. In designing these three T-DNA vectors, care was taken that (i) the selectable marker cassette neomycin phosphotransferase (NPT) is in all cases subject to similar random influences of surrounding plant DNA and (ii) no directly neighboring genes were transcribed in opposite directions. Although an obvious consequence of these requirements is that not all genes have precisely the same relative orientation with respect to the Agrobacterium T-DNA border sequences, the gene configurations resulting from integration (Figure 1) allow optimal assessment of coordinated gene expression *in planta*.

Determination of GUS and LUC enzymatic activity: For GUS and LUC assays, plant extracts were prepared by grinding 0.8-cm^2 leaf discs of greenhouse-grown plants in luciferase lysis buffer (100 mm potassium phosphate, pH 7.8; 0.2% Triton X-100; 1 mm dithiothreitol) according to the manual of the luciferase assay kit (Tropix, Bedford, MA) and cleared by centrifugation for 5 min at 4°. GUS activity was determined as described previously (Nap *et al.* 1992) and expressed as picomoles of methyl-umbelliferone per minute per microgram of soluble protein. Luciferase activity was determined in the same leaf extracts using a luciferase assay kit (Tropix) according to the recommendations of the manufacturer. Bioluminescence reactions were performed in white microtiter plates (2FB; Dynatech, Chantilly, VA) with 10 μl of plant extract and 10 μl of the luciferase assay buffer, using a Labsystems (Marlboro, MA) Luminoskan DS dual injector luminometer (van Leeuwen *et al.* 2000). The luciferase reaction was started by automatic injection of 20 μl of the substrate buffer luciferin (Tropix). Five seconds after injection the light produced during 10 sec of reaction was measured as luciferase activity. Luciferase activity was subsequently expressed as relative light units (RLU) per milligram of soluble protein. A dilution of purified luciferase (Roche, Indianapolis) was used

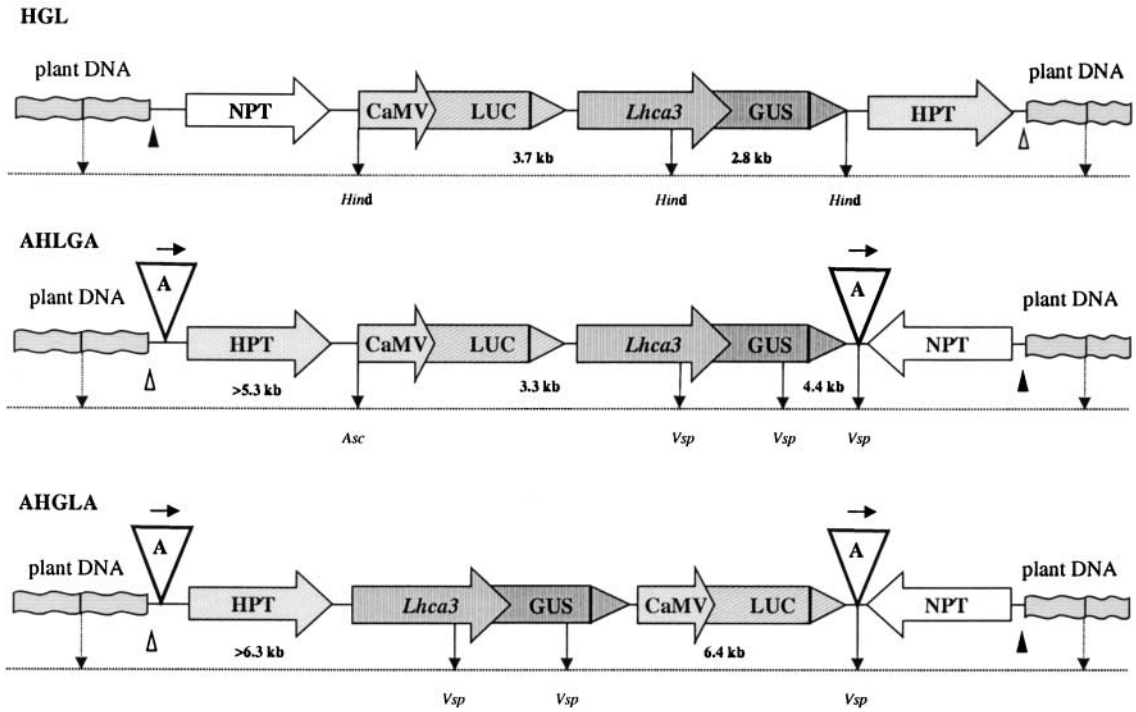

FIGURE 1.—Structure of the T-DNA region present in plants. In all three cases, the two selection genes, the hygromycin phosphotransferase gene (HPT) and the neomycin phosphotransferase gene (NPT), are driven by the *nos* promoter. The arrow gives the direction of transcription of the selection genes. The GUS gene is driven by the potato *Lhca3.St.1* promoter (*Lhca3*) and the luciferase gene (LUC) by the single 35S CaMV promoter (CaMV). The arrows give the direction of transcription by these two promoters. The chicken lysozyme matrix-associated region known as A element is represented by open triangles labeled A. The arrows above the A element give the relative orientation of the A-element DNA. These DNAs are not drawn to scale. The restriction enzymes sites used for DNA blot analysis and the expected (minimum) sizes of the relevant hybridizing fragments in plant genomic DNA are indicated. The remains of the T-DNA border sequences are indicated by solid (left border) and open (right border) triangles; curved boxes indicate flanking genomic DNA and dashed arrows indicate hypothetical restriction sites in the flanking plant DNA. *Hin*d, *Hin*dIII; *Asc*, *Asc*I; *Vsp*, *Vsp*I.

for calibration and to establish that 1 RLU per second corresponds to 25 pg of luciferase. The protein concentration in plant extracts was determined using the Bio-Rad protein assay. Data were evaluated with the analysis tools of both Excel 97 and GenStat5 (GENSTAT 5 COMMITTEE 1993).

DNA and RNA analysis: Genomic DNA was isolated from tobacco leaf material by the urea-phenol extraction procedure as described previously (MLYNÁROVÁ *et al.* 1994). Genomic DNA was digested with the appropriate restriction enzymes *Vsp*I and *Hin*dIII (Life Technologies) and *Asc*I (New England BioLabs) using the conditions recommended by the manufacturer, separated on 0.8% agarose gel and electroblotted on Hybond-N+ membrane (Amersham, Buckinghamshire, UK). Probes were isolated as restriction fragments from agarose gel and were labeled with [^{32}P]dATP using a Megaprime DNA labeling system (Amersham). Prehybridization was performed at 65° in hybridization solution (10% dextran sulfate, 1% SDS, 1 M NaCl, 0.1 mg/ml sonicated salmon sperm DNA) for at least 1 hr. Hybridization was performed overnight at 65° in a hybridization oven (Hybaid). After hybridization, the membrane was washed first in 2× SSC, 1% SDS, followed by washing in 1× SSC, 1% SDS at 65°. Hybridizing signals were visualized by autoradiography on Fuji 100NIF films and quantified using a Bas2000 PhosphorImager (Fuji) with BasReader and TINA software (Raytest). For rehybridization, blots were stripped by incubation for 15–30 min in 0.4 N NaOH, followed by 15 min incubation in 0.1× SSC, 0.2 M Tris-HCl pH 7.5 at 42° with gentle agitation and checked for the absence of any remaining signal.

RNA was isolated from leaf material using the TRIZOL reagent (Life Technologies) according to the manufacturer's recommendations. Total RNA was separated on a 1.5% formaldehyde agarose gel and vacuum blotted in 10× SSC onto a GeneScreenPlus membrane (DuPont). Prehybridization was performed at 60° in hybridization solution (10% Dextran sulfate, 1% SDS, 1 M NaCl, 0.2 mg/ml sonicated salmon sperm DNA) for at least 1 hr. Prior to hybridization, a labeled probe was boiled for 10 min in the presence of 4 mg/800 μl of sonicated salmon sperm DNA. Hybridization was performed overnight at 60° in a Hybaid hybridization oven. After hybridization, the membrane was washed first in 2× SSC, 1% SDS, followed by washing in 1× SSC, 1% SDS at 60°. For rehybridization, a blot was stripped by four to five periods of short incubation (2 min) in a boiling solution of 0.01% SDS and 0.01× SSC and checked for the absence of any remaining signal. Hybridizing signals were visualized by autoradiography on Fuji 100NIF films and quantified using a Bas2000 PhosphorImager (Fuji) with BasReader and TINA software (Raytest). Statistical analyses were performed with the analysis tools of Excel 97 as well as with GenStat5 (GENSTAT 5 COMMITTEE 1993).

RESULTS

Generation and molecular characterization of tobacco transformants: The T-DNA configurations introduced in tobacco plants are shown in Figure 1. For

TABLE 1

Characteristics of T-DNA integrations in the three populations of plants

	Population name and no. of plants in the population[a]		
Integration[b]	HGL	AHLGA	AHGLA
Total[c] (%)	37 (100)	27 (100)	33 (100)
Complex[d] (%)	28 (76)[e]	12 (44)	15 (45)
Simple[f] (%)	9 (24)	15 (56)	18 (54)
One-copy[g] (%)	5 (14)	10 (37)	9 (27)
Two-copy (%)	3 (8)	4 (15)	9 (27)
More-than-two copies (%)	1 (2)	1 (4)	0 (0)

[a] Plant populations are named after the T-DNA they carry (see Figure 1).
[b] Type of T-DNA integration as determined by DNA blot analysis.
[c] Total population of plants with at least one complete integration of the T-DNA.
[d] Population of plants with a complex integration of the T-DNA (based on an unequal number of left and right border fragments).
[e] The percentage of plants relative to the corresponding total population is given in parentheses.
[f] Population of plants with a simple integration of the T-DNA, based on the equal number of left and right border fragments.
[g] Number of plants with one, two, or more than two simple T-DNA copies integrated intact.

each, a population of transgenic plants was generated by Agrobacterium-mediated transformation and selection on kanamycin. Previous experience, as well as literature data (*e.g.*, Muskens *et al.* 2000), have shown that incomplete and complex T-DNA configurations in plants influence gene expression and complicate the subsequent analysis of gene expression data. Therefore, transformants with at least one complete integration were identified by rooting on hygromycin and extensive DNA blot analyses. Plants containing an incomplete integration of the incoming T-DNAs were omitted from further analysis. The transformants with at least one complete integration of the respective T-DNA are referred to as "total" in Table 1 and in the remainder of this article. These plants were analyzed in more detail.

To determine the number of T-DNA integrations, DNA was hybridized with an NPT and HPT probe. In addition, GUS and LUC probes were used to confirm the correct integration of these genes by the presence of hybridizing fragments of known size. In Figure 1, the predicted minimal fragment sizes of the various restriction enzyme digests of plant genomic DNA are indicated.

Plants that contained different numbers of left and right border fragments were classified as having complex integrations, such as inverted repeats or truncated integrations. These subpopulations are referred to as "complex" in Table 1 and in the remainder of this article. Complex T-DNA integrations were found in all three populations, but the relative distributions differ (Table 1). The HGL control population has most plants with such complex integration patterns: 28 plants out of 37, which is 76% of the population. In contrast, the AHGLA and AHLGA populations contain 45 and 44% plants with such complex integrations (Table 1).

Plants with equal numbers of left and right border fragments were classified as having simple and intact integrations of the whole T-DNA. These subpopulations are referred to as "simple" in Table 1. Depending on the number of border fragments, plants contain one, two, or more than two simple integrations. The class of simple integrations was therefore further subdivided into the subclasses "one-copy," "two-copy," or "more-than-two copies" (Table 1). Only 5 (14%) of the HGL population consists of one-copy transformants. The two populations carrying the A element have considerably higher percentages of one-copy integrations than the HGL population (Table 1). For the one-copy plants, the single copy of the T-DNA was confirmed by genetic analysis of selfed seeds showing a 3:1 segregation on kanamycin as well as hygromycin. Most two-copy transformants analyzed showed a 15:1 segregation on kanamycin and hygromycin, indicating the presence of two unlinked T-DNAs (data not shown).

Analyses of GUS and LUC enzymatic activities in the total populations: GUS and LUC enzymatic activities were determined in comparable leaf samples from the total populations of plants (Table 1). The GUS enzymatic activity is expressed in picomoles of methylumbelliferone per minute per microgram of soluble protein. The LUC activity is expressed in relative light units per milligram of soluble protein. As before, proper statistical analyses required a logarithmic transformation for both the GUS and the LUC activity data (analysis not shown; Nap *et al.* 1993b). In Figure 2, A–C, the natural logarithm (ln) of the GUS activity is shown for each transformant in each of the populations. In Figure 2, D–F, the natural logarithm of the LUC activity in each plant is plotted. In these plots, the total population is represented by the sum of the blue, red, and open circles.

In Table 2, the descriptive statistics are given for these

six data sets. Comparing the mean activity values, these data suggest that the gene present in the middle of the MAR-delimited loop tends to have a higher activity than when the same gene is closer to the A element. The AHGLA population, with the GUS gene in the middle, has an average GUS activity of 278.6 pmol methylumbelliferone per minute per microgram of protein, whereas the AHLGA population has an activity of only 148.4, almost half that of the AHGLA population. For the LUC gene this trend is reversed.

Comparing variances, the MAR-containing populations show a significantly reduced variation in GUS activities compared to the control population HGL. The AHLGA and AHGLA populations show 4.3- and 3.5-fold reduction of variance compared to HGL, which is highly significant (Table 2). As shown before (Mlynárová *et al.* 1994, 1995), the presence of the A element is accompanied by a significant reduction of the variation of GUS activity between independent transformed plants. In contrast, no significant differences in variation were observed for the LUC activity data compared to the control population (Table 2), suggesting that the MAR elements have had no detectable influence on the variation in LUC activity. Further analyses (see below) show that this result is due partly to the presence of complex integrations in this total population.

Of special interest in this study is the relation between the activities of the two genes present on the composite T-DNA. Figure 2, G–I, shows the double logarithmic plot of GUS activity for each transformant plotted against the LUC activity of the same transformant. Table 3 gives the product moment correlation coefficients (R) for these associations. In the control population HGL, there is no apparent correlation between the activities of both genes ($R = 0.24$), whereas the relationship between both activities is markedly increased in the AHGLA ($R = 0.83$) and AHLGA ($R = 0.81$) populations. Spearman's rank correlation, which is not dependent on the logarithmic transformation procedure, yields the same conclusions (not shown). These results indicate that for the MAR-containing populations, plants with high GUS activity also show high LUC activity and vice versa. Part of the improved correlation in the MAR-containing populations could be due to a copy number effect. To remove such a copy number effect from the plot, the ratio of logarithmic GUS and LUC activity [*i.e.*, ln(GUS)/ln(LUC)] is plotted in Figure 2, J–L. The nearly straight line obtained in the case of the AHGLA and AHLGA populations shows that GUS and LUC activity in all the MAR-containing plants yields virtually the same ratio, whereas the ratio in the HGL control population is much more variable. These data are quantified and analyzed in Table 4. The data in Table 4 illustrate that the variance of the ln(GUS)/ln(LUC) ratio of the AHGLA and AHLGA populations is highly significantly reduced compared to the control population HGL. The activity levels of the two genes become more associated, supposedly due to a common cause, *i.e.*, the presence of the MAR elements.

Plants with complex integrations may show post-transcriptional gene silencing (Muskens *et al.* 2000) and tend to have more unstable and less predictable transgene expression levels. Such plants can influence the statistical evaluation of populations considerably. In Table 2, the descriptive statistics of all plants classified as complex are given. In Figure 2, A–L, the complex plants are plotted as open circles. Comparing the total populations with the complex subsets, it can be concluded that there is little difference (Tables 2–4). On the basis of correlation coefficients (Table 3) as well as on ln(GUS)/ln(LUC) variances (Table 4), the two MAR-containing populations behave more consistently than the complex subpopulation from the HGL control.

Analyses of GUS and LUC enzymatic activities in plants with simple T-DNA integrations: To eliminate the potential influence of complex T-DNA integrations, the quantitative analyses were restricted to those plants that had simple integration patterns. This reduces the total number of plants available for analysis to 42. Table 2 gives the descriptive statistics of the data sets of these plants. In Figure 2, the simple plants are indicated by the sum of the red and the blue circles. In Figure 2, A–C, the GUS activity data are plotted and in Figure 2, D–F, the LUC activity data, respectively. Figure 2, G–I, shows the double logarithmic plot of GUS activity against the LUC activity and Figure 2, J–L, the copy-number-corrected ln(GUS)/ln(LUC) ratio. Table 3 gives the GUS/LUC correlation and Table 4 gives the mean and variance of the ratios plotted in Figure 2, J–L. For the GUS activity, the two MAR-containing populations still show a significantly reduced variation compared to the control population HGL. Removing the complex integrations from the analysis now reveals a 4.8-fold significant reduction in variance for LUC activity in the AHLGA population compared to the control population HGL (Table 2). The variance in the AHGLA luciferase activity is ~2-fold lower than the variance in the HGL population, which is barely not significant (at $P = 0.05$; Table 2) due to copy number effects (Table 4). When copy number effects are eliminated (Table 4), the AHGLA simple population also shows a significant reduction in the variation of LUC activity. The correlation coefficients (Table 3) follow the trends given by the total population: in the control population HGL, there is still no correlation between the activities of both genes ($R = 0.31$), whereas the relationship is obvious in the MAR-containing AHGLA ($R = 0.80$) and AHLGA ($R = 0.73$) populations.

In plants with multiple simple integrations of transgenes, these transgenes may (epigenetically) interact and deviate from the expected copy-number-dependent additivity of gene expression. Therefore, a final analysis was based on all plants with a simple one-copy integration of the T-DNA. The total number of plants available for this analysis is 24. Table 2 gives the descriptive statis-

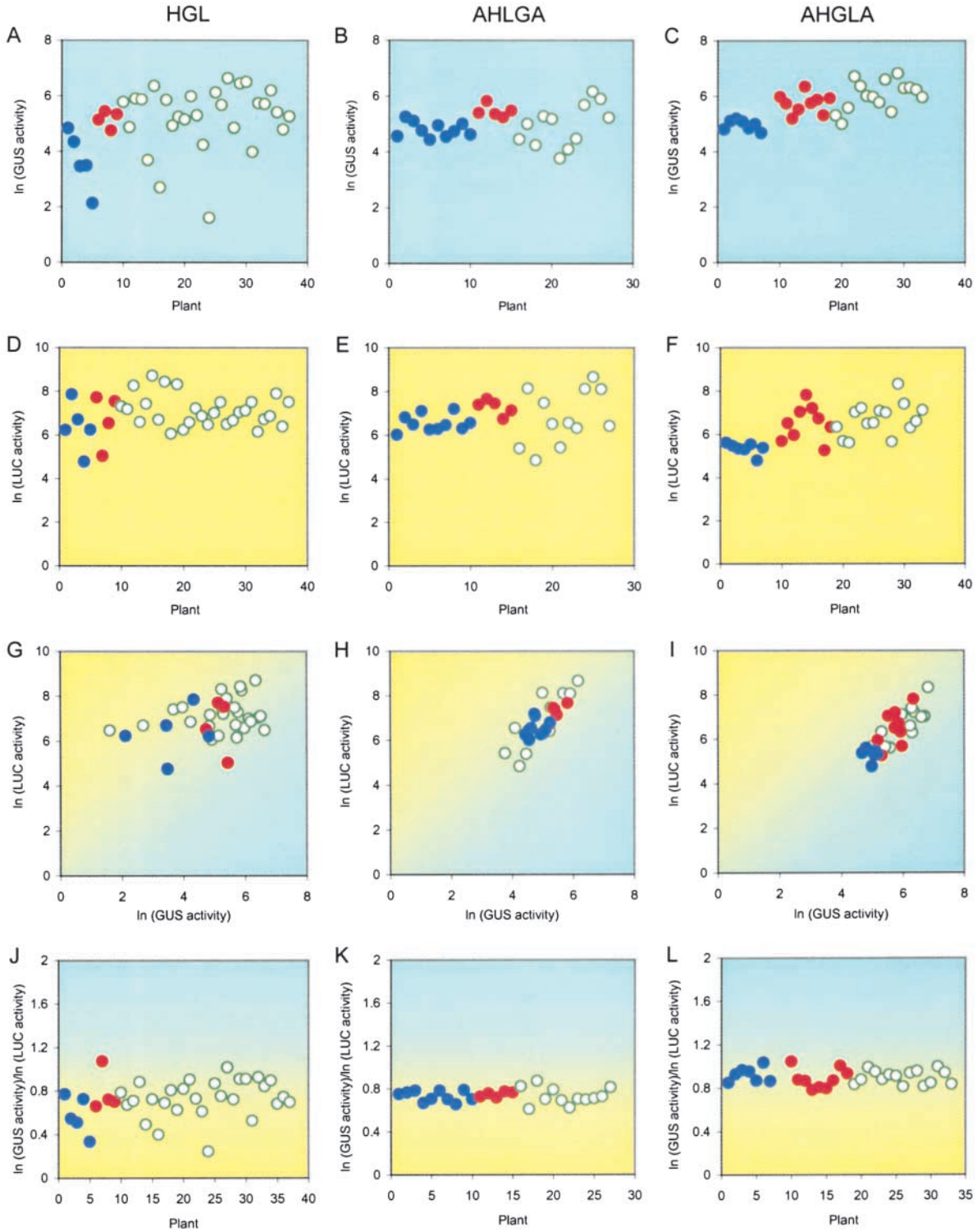

FIGURE 2.—Quantified enzymatic activities in all transformed plants in the three populations. The population name is given at the top. (A–C) The ln of the GUS activity. (D–F) The natural logarithm of the luciferase activity. (G–I) The double logarithmic plot of GUS activity plotted against the LUC activity of the same transformant. (J–L) The ratio of the logarithmic GUS and LUC activity. In A–L, plants with a one-copy integration of the T-DNA are indicated in blue, plants with two or more simple T-DNA integrations are in red, and all other plants with complex integrations are represented by open circles. Panels related to the GUS gene are in blue, and those related to the LUC gene are in yellow. Panels combining the two genes are in a combination of both colors.

TABLE 2

Descriptive statistics of the enzymatic activities in the three populations of plants

		GUS activity					LUC activity			
Integration[a]	Population[b]	No.[c]	Mean[d]	Var.[e]	Fold[f]	*P*[g]	Mean	Var.	Fold	*P*
Total	HGL	37	5.01	1.43	—[h]	NA[i]	6.98	0.72	—	NA
	AHLGA	27	5.00	0.33	4.3	***[j]	6.81	0.80	0.90	NS
	AHGLA	33	5.63	0.41	3.5	***	6.27	0.77	0.93	NS
Complex	HGL	28	5.24	1.33	—	NA	7.13	0.50	—	NA
	AHLGA	12	4.96	0.57	2.3	NS∼	6.84	1.57	0.32	**
	AHGLA	15	6.02	0.28	4.7	**	6.7	0.55	0.91	NS
Simple	HGL	9	4.32	1.20	—	NA	6.52	1.21	—	NA
	AHLGA	15	5.02	0.16	7.5	**	6.80	0.25	4.84	**
	AHGLA	18	5.30	0.29	4.1	**	5.92	0.70	1.73	NS
One-copy	HGL	5	3.64	1.06	—	NA	6.37	1.23	—	NA
	AHLGA	10	4.69	0.08	13.2	***	6.55	0.14	8.8	**
	AHGLA	9	4.75	0.06	18.4	***	5.33	0.08	15.4	***
Two-copy	HGL	3	5.10	0.12	—	NA	6.44	1.79	—	NA
	AHLGA	4	5.45	0.06	2	NS	7.32	0.16	11.2	*
	AHGLA	9	5.73	0.12	1	NS	6.51	0.64	2.8	NS∼

[a] Type of T-DNA integration as determined by DNA blot analysis (see text and Table 1); total is complex plus simple; simple is further subdivided into one-copy and two-copy plants.
[b] Populations are named according to the T-DNA construct the plants contain (see Figure 1).
[c] Number of plants in the population.
[d] Means are based on the natural logarithms of enzyme activities.
[e] Variance.
[f] Fold reduction of variance with respect to the corresponding control population HGL.
[g] One-tailed probability according to the *F*-test for homogeneity of variances in comparison to the corresponding control population HGL.
[h] (—) Control population HGL.
[i] Not applicable.
[j] Indication of the probability value at the 95% confidence level; ***, significant at $P \leq 0.001$; **, significant at $P \leq 0.01$; *, significant at $P \leq 0.05$; NS∼, barely not significant ($0.05 < P \leq 0.1$); NS, not significant (at $P > 0.1$).

tics of the one-copy populations. In Figure 2, the one-copy plants are indicated by the blue circles. In Figure 2, the GUS activity data are plotted in A–C and the LUC activity data are plotted in D–F. In Figure 2, G–I, the double logarithmic plot of GUS activity against the LUC activity is given and in Figure 2, J–L, the derived ln(GUS)/ln(LUC) ratio is given. Table 4 gives the mean and variance of the ratios plotted in Figure 2, J–L. For both GUS and LUC the variation in one-copy AHGLA and AHLGA plants is reduced significantly ($P < 0.001$), up to 18-fold, compared to the HGL population (Table 2), as is also apparent from Figure 2, A–C (for GUS) and Figure 2, D–F (for LUC). Both the one-copy AHGLA and the AHLGA populations cluster together in double logarithmic plots (Figure 2, H and I, blue circles). In contrast, the HGL data show a considerable scatter (Figure 2G, blue circles), despite the low number of only 5 plants present in this population. The correlation coefficient is not a very suitable way of demonstrating a clustering of plants toward a single point. Due to the small population sizes, none are significantly different from zero (data not shown). In the case of the non-MAR HGL one-copy plants, the covariance of GUS and LUC enzymatic activity is much higher than for the MAR-containing one-copy plants (data not shown). The difference between the HGL, AHGLA, and AHLGA one-copy populations is also characterized by a significant reduction in the variance of the ln(GUS)/ln(LUC) ratio (Figure 2, J–L; Table 4). This shows the highly coordinated expression of the GUS and LUC activities in the MAR-containing one-copy populations.

Relationship between GUS and LUC enzymatic activities and mRNA levels: The data demonstrating the efficacy of the chicken lysozyme A element in the reduction of position effects are based on enzymatic activities. This implicitly assumes linear relationships between transcription rate, steady-state mRNA accumulation, amount of protein, and the enzymatic activity of that protein. One of the concerns is whether and to what extent the quantitative activity data reflect the steady-state RNA accumulation of the genes used. Therefore, the analyses were extended to the RNA levels of GUS and LUC in the three populations of transgenic plants. The remainder of almost every leaf from which a sample was taken

TABLE 3

Correlation of GUS and luciferase expression data

		Compared parameters[a]											
		GUS activity-LUC activity			GUS activity-GUS mRNA			LUC activity-LUC mRNA			GUS mRNA-LUC mRNA		
Integration[b]	Population[c]	No.[d]	*R*[e]	*P*[f]	No.	*R*	*P*	No.	*R*	*P*	No.	*R*	*P*
Total	HGL	37	*0.24*	NA[g]	33	0.55	NA	33	*0.07*	NA	33	0.80	NA
	AHLGA	27	0.81	***[h]	27	0.75	NS∼	27	0.54	*	27	0.91	*
	AHGLA	33	0.83	***	30	0.80	*	30	0.53	*	30	0.90	NS∼
Complex	HGL	28	*0.23*	NA	25	0.57	NA	25	*0.04*	NA	25	0.83	NA
	AHLGA	12	0.86	***	12	0.85	NS∼	12	0.63	*	12	0.90	NS
	AHGLA	15	0.79	***	14	0.82	NS∼	14	*0.38*	NS	14	0.81	NS
Simple	HGL	9	*0.31*	NA	8	*0.56*	NA	8	*0.07*	NA	8	0.73	NA
	AHLGA	15	0.73	**	15	0.66	NS	15	0.56	NS	15	0.95	*
	AHGLA	18	0.80	**	16	0.67	NS	16	0.46	NS	16	0.94	NS∼

[a] Comparisons are based on the natural logarithms of the data on enzymatic activity and mRNA amount.
[b] Type of T-DNA integration as determined by DNA blot analysis (see text and Table 1); total is complex plus simple.
[c] Populations are named according to the T-DNA construct that the plants contain (see Figure 1).
[d] Number of combined data points.
[e] Product-moment (*i.e.*, the common) correlation coefficient corrected for the bias in the case of small samples from bivariate normal distributions (Sokal and Rohlf 1995). The correlation coefficients given in italics are not significantly different from zero at $P = 0.05$ according to the *t*-test for significance.
[f] One-tailed probability according to the Fisher *z* transformation with *t* approximation (Zar 1996) for equality of the correlation coefficient in comparison to the corresponding control population HGL.
[g] Not applicable.
[h] Indication of the probability value at the 95% confidence level; ***, significant at $P \leq 0.001$; **, significant at $P \leq 0.01$; *, significant at $P \leq 0.05$; NS∼, barely not significant ($0.05 < P \leq 0.1$); NS, not significant (at $P > 0.1$).

for the determination of the enzymatic GUS and LUC activities was used for total RNA isolation. The relative amounts of GUS and LUC transcripts as well as 18S rRNA amounts were determined with radioactively labeled probes on RNA blots. In Figure 3, representative examples of the RNA blot analyses for each of the three populations are shown. Data were quantified with the help of a PhosphorImager. The GUS and LUC RNA signals were normalized for the amount of RNA loaded relative to the rRNA signal. Also for the normalized RNA amounts, a natural logarithmic transformation was appropriate for statistical evaluation (analysis not shown), so all subsequent analyses were performed after such a transformation.

In Figure 4, the analysis of the RNA data for the total populations is shown. As in Figure 2, this is the sum of the red, blue, and open circles. The relationship between GUS activity and amount of GUS mRNA is plotted in Figure 4, A–C, and the relationship between LUC activity and amount of LUC mRNA is plotted in Figure 4, D–F. The corresponding correlations are given in Table 3. Whereas there is no correlation between LUC mRNA amount and LUC activity in the HGL population ($R = 0.07$), the equivalent correlation in the MAR-containing plants is still poor but markedly improved ($R = 0.54/0.53$) and significantly different from zero. For GUS, the same trend is apparent: the HGL population shows a rather poor correlation ($R = 0.55$), whereas the AHLGA and AHGLA populations are improved ($R = 0.75/0.80$). These data indicate, therefore, that the overall correlation between enzymatic activity and mRNA amount is disappointing. The presence of the MAR elements helps to improve that correlation. In Figure 4, G–I, the GUS mRNA is plotted against the LUC mRNA amount; the corresponding correlation coefficients are given in Table 3. Despite the poor correlations between enzymatic activity and mRNA, the correlations between GUS and LUC mRNA are high. As shown above for the activity data, for the mRNA amount data a major part of the correlation also is likely to be due to copy number effects. Figure 4, J–L, shows the plot of the ln(GUS mRNA)/ln(LUC mRNA) ratio and the accompanying statistics are given in Table 4. Compared to the control population HGL, a reduction in the ln(GUS mRNA)/ln(LUC mRNA) ratio is observed. Therefore, the expression of both genes has become coordinated on the level of steady-state RNA amount as well. Similar analyses of the subset of plants with simple copies (sum of red and blue circles, Figure 4, A–L; Tables 3 and 4) and the simple one-copy plants (blue circles, Figure 4, A–L) further substantiate the results obtained for the total population. Due to the relatively small population sizes, the confidence limits of the statistical parameters estimated are large, so only trends can

TABLE 4

Descriptive statistics of the enzymatic and mRNA GUS/LUC ratios

		ln(GUS activity)/ln(LUC activity)					ln(GUS mRNA)/ln(LUC mRNA)				
Integration[a]	Population[b]	No.[c]	Mean[d]	Var.[e]	F[f]	P[g]	No.	Mean	Var.	F	P
Total	HGL	37	0.72	0.032	NA[h]	NA	32	0.86	0.027	NA	NA
	AHLGA	27	0.74	0.004	8.0	***[i]	27	0.95	0.014	1.93	NS∼
	AHGLA	33	0.90	0.005	6.4	***	30	0.91	0.014	1.93	*
Complex	HGL	28	0.74	0.029	NA	NA	25	0.88	0.029	NA	NA
	AHLGA	12	0.73	0.006	4.8	**	12	0.96	0.024	1.2	NS
	AHGLA	15	0.90	0.004	7.2	***	14	0.94	0.018	1.6	NS
Simple	HGL	9	0.66	0.040	NA	NA	8	0.78	0.014	NA	NA
	AHLGA	15	0.74	0.002	20	***	15	0.94	0.007	2.0	NS
	AHGLA	18	0.90	0.006	6.7	***	16	0.89	0.009	1.4	NS
One-copy	HGL	5	0.58	0.030	NA	NA	4	0.70	0.009	NA	NA
	AHLGA	10	0.73	0.003	10	***	10	0.92	0.009	1	NS
	AHGLA	9	0.91	0.004	7.5	**	7	0.83	0.013	0.7	NS

[a] Type of T-DNA integration as determined by DNA blot analysis (see text and Table 1); total is complex plus simple; one-copy is a further subdivision of simple.
[b] Populations are named according to the T-DNA construct the plants contain (see Figure 1).
[c] Number of plants used for analysis.
[d] Means are based on the natural logarithms of enzyme activities or RNA amounts.
[e] Variance.
[f] Fold reduction of variance with respect to the corresponding control population HGL.
[g] One-tailed probability according to the F-test for homogeneity of variances in comparison to the corresponding control population HGL.
[h] Not applicable.
[i] Indication of probability value at the 95% confidence level; ***, significant at $P \leq 0.001$; **, significant at $P \leq 0.01$; *, significant at $P \leq 0.05$; NS∼, barely not significant ($0.05 < P \leq 0.1$); NS, not significant (at $P > 0.1$).

be pointed out. The relationship between GUS mRNA and GUS activity, on the one hand, and LUC mRNA and LUC activity, on the other hand, improve in the two MAR-containing populations compared to the HGL control.

DISCUSSION

The expression of two different genes placed between chicken lysozyme A-element DNA becomes highly correlated and coordinated, in contrast to the situation where the genes are flanked by only selectable marker genes. Both enzymatic activity levels and steady-state RNA levels of the two genes become coordinated, suggesting that the influence of the A elements is on the level of transcription. The artificial chromatin domain created by the chicken lysozyme A element is apparently establishing a functional unit of gene regulation in plants.

Whereas the phenotypic effects observed in the presence of the A elements are obvious, the molecular mechanism of this biological activity of the elements is a controversial issue. It is an open question whether the observed coordination is due to nuclear matrix binding or to some other specific, but yet unknown, characteristic of the A element. The presence of the A element on transforming DNA could, for example, somehow help to target the transgenes to active regions of the genome (Allen *et al.* 2000). In different laboratories, different results have been obtained with different MAR elements, genes, and recipient cells or species, suggesting that results cannot easily be generalized. Our control (see Figure 1) is devised in such a way that the two selectable marker genes (NPT and HPT) surround the reporter genes (GUS and LUC). Both marker genes are active in the plants analyzed. These genes are evaluated as an alternative to the A element in shielding transgenes from the influences of neighboring chromatin. Bhattacharyya *et al.* (1994) showed a reduced GUS variability using a T-DNA with two selectable markers located at the T-DNA borders, similar to the transgene configuration analyzed here. Compared to previous results (Mlynárová *et al.* 1994), the observed reduction in variability is not merely a matter of selectable marker DNA around the gene(s) of interest. A reduction in variability requires the presence of DNA with the characteristics of the A element. Besides its presumed role in nuclear architecture, the A element could collaborate with enhancer and/or promoter sequences (Forrester *et al.* 1994), generate an extended domain of open chromatin (Zhao *et al.* 1993; Jenuwein *et al.* 1997), or influence transcription and pre-mRNA processing components (Göhring and Fackelmayer 1997; Nayler *et al.* 1998).

Another issue of debate is the role of the configura-

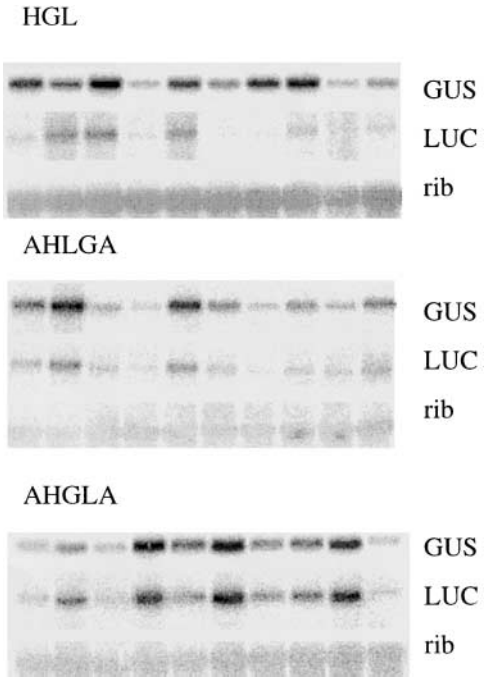

FIGURE 3.—Autoradiograph of a RNA blot showing the RNA amounts in a random subset of plants from each of the three populations. Total RNA was separated on formaldehyde gels, blotted, probed successively with labeled probes for the GUS gene, the LUC gene, and the pea 18S ribosomal sequences (rib), and visualized by phosphorimaging. The phosphor images shown have had different exposure times that were selected to give approximately comparable signals.

tion of genes within the A-element-delimited DNA loop. The two test constructs evaluated here (see Figure 1) were devised in such a way that they do not have an A element immediately upstream of the promoter of either the GUS or the LUC reporter gene. This was based on our previous results (MLYNÁROVÁ *et al.* 1994, 1995) and follows suggestions that transcription may be hampered by too close an association with the nuclear matrix (RAZIN *et al.* 1995; STRÄTLING and YU 1999). Moreover, the configuration of a *nos* promoter relative to the dCaMV and *Lhca3.St.1* promoter-gene fusions is kept the same to eliminate potential problems with any putative readthrough. The analyses show that the LUC gene yields most activity in the AHLGA population when in the middle of the chromatin loop. The GUS gene shows the same trend. When in the middle of the presumed loop of A-element-delimited DNA, the gene tends to be most active. This indicates that a gene may be most accessible for the transcription machinery when placed away from the A elements.

The variation in GUS activity in the MAR-containing populations analyzed here is significantly lower than in the non-MAR control analyzed previously (MLYNÁROVÁ *et al.* 1994). Therefore, the MAR elements also reduce the position effect when the presumed DNA loop between the two elements is enlarged. A reduction in the variation of LUC activity could be demonstrated only in the subpopulations of one-copy plants (Table 2). No differences in variance of LUC enzymatic activities were observed between the MAR-containing population and the HGL control population. A possible explanation is that the CaMV promoter driving the LUC gene is more prone to epigenetic effects than the *Lhca3.St.1* promoter driving the GUS gene (MLYNÁROVÁ *et al.* 1996). In one-copy plants, the MAR elements reduce the position effect in the luciferase activity as much as they reduce the position effect of GUS (GUS: 13- and 18-fold reduction; LUC: 9- and 15-fold reduction). Therefore, in this subpopulation the MAR elements also do protect the luciferase gene from the influences of the surrounding chromatin and this reduction of position effects reveals the correlation between the activity levels.

In subsets of the populations defined on the basis of T-DNA configuration, the differences in variance between MAR and non-MAR plants become larger. This confirms that complex integrations and copy number dependence of gene activity influence the variance in the populations of transgenic plants. Complex integrations are often associated with gene-silencing phenomena (MUSKENS *et al.* 2000). Comparisons of the three complex populations generated here show that the presence of MAR elements improves the expression characteristics in terms of variance reductions and correlations (Tables 2–4). The presence of MAR elements may therefore protect (in part) against transcriptional silencing induced by *cis* interactions within complex and/or repeated transgene arrays (ALLEN *et al.* 2000). A copy number effect resulting in additive enzyme activities of multiple integrations obviously increases the variance when the total population is evaluated. The moment populations differ in the relative proportion of complex and multicopy integration events, comparisons and conclusions about MAR action may be erroneous or uninformative. The differences between the GUS and LUC data in the total populations are likely to be due partly to the characteristics of the LUC protein. Obviously, both mRNA and protein stability play an important role in establishing both the level and the kinetics of gene expression. The LUC protein is degraded rapidly, notably in the absence of its substrate (THOMPSON *et al.* 1991). As a result, it provides a better and instantaneous monitor of gene expression, but will also be more variable than a long(er)-living reporter molecule such as GUS.

The data show an unexpected and surprisingly poor correlation between enzymatic activity and RNA amount (Table 3) for both GUS and LUC. In large-scale comparisons in yeast of the transcriptome with the proteome, only the high abundant proteins showed a good correlation between mRNA amount and protein. Out of 106 genes analyzed, no less than 95 had a correlation of 0.5 or less, the majority showing a correlation of <0.3 (GYGI *et al.* 1999). Generally, lack of correlation between RNA and protein levels is thought to be due to post-transla-

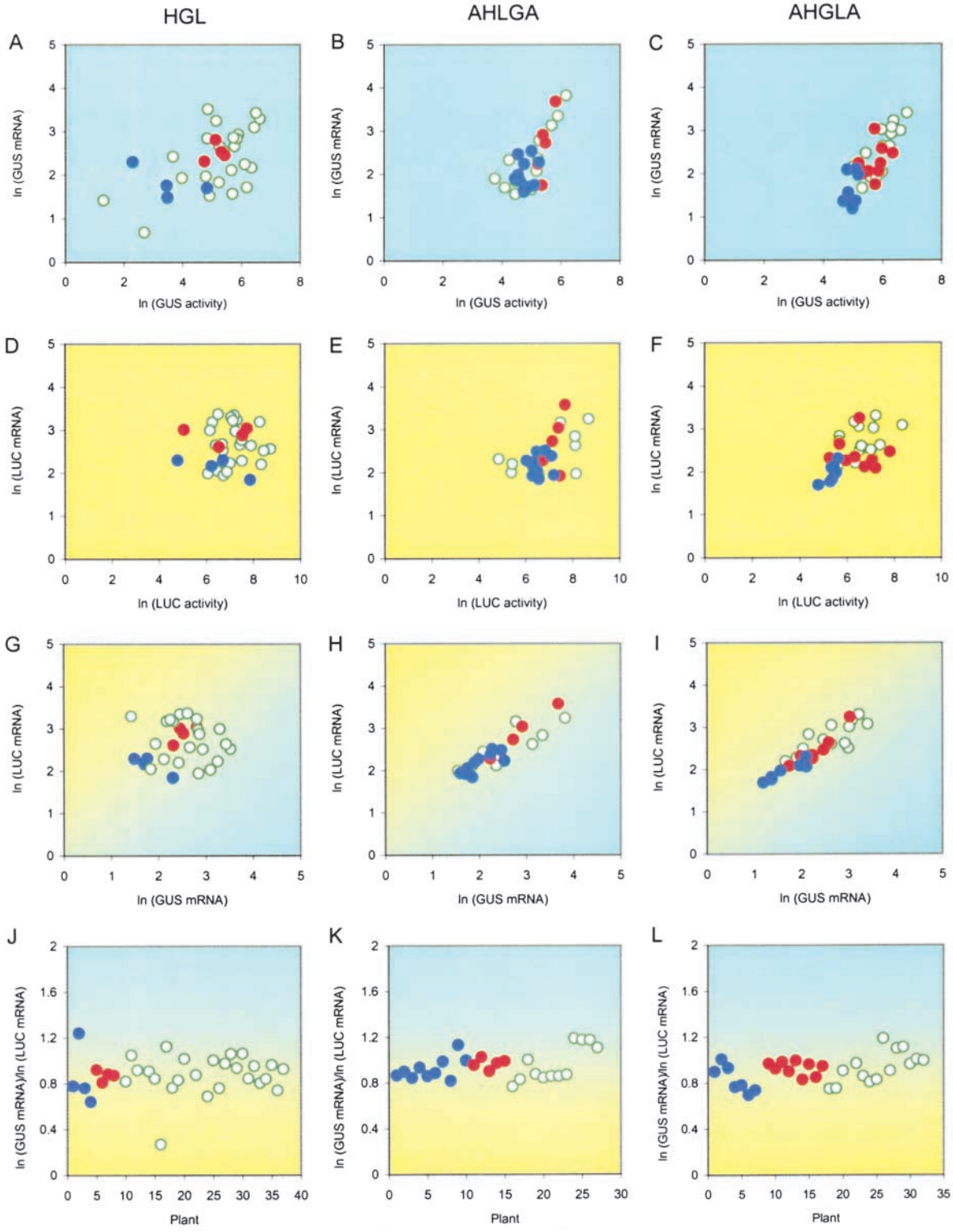

FIGURE 4.—Quantified GUS and LUC mRNA amounts in all transformed plants in the three populations. Total RNA was analyzed as indicated in Figure 3. The relative amounts of RNA transcripts were quantified with a PhosphorImager; the GUS and LUC signals were normalized for the amount of RNA loaded relative to the hybridization with the ribosomal probe. (A–C) The double logarithmic plot of GUS activity and normalized GUS mRNA. (D–F) The natural logarithm of LUC activity and normalized LUC mRNA. (G–I) The double logarithmic plot of the amount of GUS mRNA plotted against the LUC mRNA in the same transformant. (J–L) The ratio of the logarithmic GUS and LUC mRNA amounts. In all panels, plants with a one-copy integration of the T-DNA are indicated in blue, plants with two or more simple T-DNA integrations are in red, and all other plants with complex integrations are represented with open circles. Panels related to the GUS gene are in blue, and those related to the LUC gene are in yellow. Panels combining the two genes are in a combination of both colors.

tional modifications and/or relative differences between mRNA and protein turnover. The latter may explain in part the difference between the GUS and LUC data. The improved correlation upon the presence of MAR elements suggests that the positive influence of the MAR elements extends beyond their role in the nucleus. They may somehow influence the export from the nucleus, RNA turnover, or possibly even the translational efficiency of the mRNA in the cytoplasm. The relatively poor correlation between RNA and enzyme activity levels creates the following situation: for the MAR-containing populations, GUS and LUC enzymatic activities are well correlated and GUS and LUC mRNA levels are well correlated, but GUS and notably LUC activities are not well correlated with their respective mRNA levels. At first, this is somewhat counterintuitive. A possible statistical explanation is the following: the experimental procedures of protein isolation and kinetic analyses, on the one hand, and RNA analysis, on the other hand, each involve a number of experimental steps, each with its own errors. Some of these errors will not be completely independent from one sample to the other, but rather work in the same direction. For example, the same isolate and protein determination is used for both the GUS and the LUC kinetic measurements. Similar arguments hold for the determination of RNA levels. When performing correlation analyses (Table 3) or when taking the activity ratio as the parameter of analysis, such as in Figure 2, J–L, and in Figure 4, J–L, all method-dependent errors cancel out and the result will be a more accurate, and therefore a more relevant, estimate. However, when enzymatic activities are combined with RNA levels, the methodological errors become much more independent. They will not cancel out, but in contrast give rise to more "noise." As a result of this increased noise, correlations can become disguised and/or undetectable. For the data presented here, the consequence of such "error pipelines" is that the correlations between enzymatic activity levels, on the one hand, and RNA levels, on the other hand, both showing coordinated expression of the genes investigated, are more informative than the poor correlation between enzymatic activity and RNA level. The concept of data generation-dependent error pipelines would seem to need more attention in the bioinformatics of the analyses of different types of data, *e.g.*, in whole-genome data sets.

Over the years, there have been numerous investigations into the possibilities and mechanisms of coordinated expression in plants. Dean *et al.* (1988a,b) showed a reasonable coordinated expression, quantified as RNA, of the chloramphenicol acetyl transferase and octopine synthase genes, when flanked with large similar, but not identical, genomic regions of well-expressed petunia *rbcS* gene regions and *rbcS* promoters. This may have been due to the presence of MAR-like sequences in those regions of petunia DNA. Other attempts with multiple genes have been less successful. Gidoni *et al.* (1988) could achieve the coordinated transgene expression only by maintaining the normal linkage of petunia divergently expressing *Cab22L* and *Cab22R* genes. The divergent *mas1,2* promoter did not yield a correlated expression of two genes (Peach and Velten 1991), nor did the simultaneous introduction of two genes on a single plasmid into tobacco by particle bombardment (Leech *et al.* 1998). For complex metabolic or regulatory pathways, multiple genes are generally required to study or achieve the desired alteration(s). In a transgenic setup, this requires sequential genetic modification (Krohn *et al.* 1998), stacking by successive crossing (Poirier *et al.* 1992, 2000; Nawrath *et al.* 1994; Ma *et al.* 1995), assembling the genes of interest on a single transforming molecule (van Engelen *et al.* 1994; Jach *et al.* 1995; Leech *et al.* 1998), or cotransformation with multiple plasmids (Chen *et al.* 1998; Maqbool and Christou 1999). Results show that when genes are combined, expression levels are highly variable and, when analyzed, there is poor correlation of the expression of cointroduced genes (Jach *et al.* 1995; Leech *et al.* 1998; Maqbool and Christou 1999; Ye *et al.* 2000). Alternative approaches using a bicistronic transcriptional unit (Iida *et al.* 1992; Lough *et al.* 1997), a self-cleaving polyprotein (Marcos and Beachy 1994, 1997; Dasgupta *et al.* 1998; Halpin *et al.* 1999), or a bifunctional protein (Elmayan and Tepfer 1994) seem more specific and less versatile than the chromatin-structure-based approach demonstrated here. For any strategy and for each gene, the best performing plant must be selected, characterized, proven sufficiently stable over time and over generations, and be compatible with the gene(s) with which it must be combined. The results presented here indicate that the use of a chromatin boundary element such as the chicken lysozyme A element may contribute to establishing a coordinated, high-level, and stable expression of multiple introduced genes. Notably the coordinated expression of two different promoter-gene combinations may have useful applications. It may allow the creation of a reference gene that is easily assayed for and functions as a "pointer" or internal standard for the transgene of interest that cannot be assayed for so easily. For example, the analysis of a gene desired to perform in the fruit of a woody plant species that takes several years to set its first fruit now requires the long waiting for that first fruit. With a proper and coordinately expressed reference gene already in the seedling stage, a selection can be made of trees that have a considerably higher likelihood of the gene of interest doing well in the fruit. The overall, significantly higher predictability of gene action when embedded in A-element DNA would make such strategies feasible and attractive.

The authors thank Alexander van der Krol and Wessel van Leeuwen (Plant Physiology Wageningen UR) for the LUC gene and fruitful discussions on the topic, Andy Pereira for plasmid pPCV720, Arjen

Jonas (Wageningen UR, Toxicology) for indispensable help with the *in vitro* luciferase measurements, the Plant Research International Greenhouse team for plant care, and Lonneke van der Geest for useful suggestions to improve the manuscript. The research was supported by the Life Sciences Foundation (SLW), which is subsidized by the Netherlands Organization for Scientific Research (NWO) as well as by program subsidy 347 from the Dutch Ministry of Agriculture, Nature Management and Fisheries.

LITERATURE CITED

Allen, G. C., G. E. Hall, Jr., L. C. Childs, A. K. Weissinger, S. Spiker *et al.*, 1993 Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. Plant Cell **5:** 603–613.

Allen, G. C., G. E. Hall, Jr., S. Michalowski, W. Newman, S. Spiker *et al.*, 1996 High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. Plant Cell **8:** 899–913.

Allen, G. C., S. Spiker and W. F. Thompson, 2000 Use of matrix attachment regions (MARs) to minimize transgene silencing. Plant Mol. Biol. **43:** 361–376.

Bhattacharyya, M. K., B. A. Stermer and R. A. Dixon, 1994 Reduced variation in transgene expression from a binary vector with selectable markers at the right and left T-DNA borders. Plant J. **6:** 957–968.

Bode, J., M. Stengert-Iber, V. Kay, T. Schlake and A. Dietz-Pfeilstetter, 1996 Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. Crit. Rev. Euc. Gene Exp. **6:** 115–138.

Breyne, P., G. Gheysen, A. Jacobs, M. Van Montagu and A. Depicker, 1992 Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. Plant Cell **4:** 463–471.

Chen, L., P. Marmey, N. J. Taylor, J.-P. Brizard, C. Espinoza *et al.*, 1998 Expression and inheritance of multiple transgenes in rice plants. Nat. Biotech. **16:** 1060–1064.

Cohen, B. A., R. D. Mitra, J. D. Hughes and G. M. Church, 2000 A computational analysis of whole genome expression data reveals chromosomal domains of gene expression. Nat. Genet. **26:** 183–186.

Conner, A. J., L. Mlynárová and J.-P. Nap, 1998 Meiotic stability of transgene expression is not affected by flanking transgenes with matrix-associated region repeats. Mol. Breed. **4:** 47–58.

Dasgupta, S., G. B. Collins and A. G. Hunt, 1998 Co-ordinated expression of multiple enzymes in different subcellular compartments in plants. Plant J. **16:** 107–116.

Dean, C., M. Favreau, S. Tamaki, D. Bond-Nutter, P. Dunsmuir *et al.*, 1988a Expression of tandem gene fusions in transgenic tobacco plants. Nucleic Acids Res. **16:** 7601–7617.

Dean, C., J. Jones, M. Favreau, P. Dunsmuir and J. Bedbrook, 1988b Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. Nucleic Acids Res. **16:** 9267–9283.

Dillon, N., and P. Sabbattini, 2000 Functional gene expression domains: defining the functional unit of eucaryotic gene regulation. Bioessays **22:** 657–665.

Elmayan, T., and M. Tepfer, 1994 Synthesis of a bifunctional metallothionein/beta-glucuronidase fusion protein in transgenic tobacco plants as a means of reducing leaf cadmium levels. Plant J. **6:** 433–440.

Forrester, W. C., C. Van Genderen, T. Jenuwein and R. Grosschedl, 1994 Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. Science **265:** 1221–1222.

Genstat 5 Committee, 1993 *Genstat 5 (Release 3) Reference Manual.* Clarendon Press, Oxford.

Gidoni, D., D. Bond-Nutter, P. Brosio, J. Jones, J. Bedbrook *et al.*, 1988 Coordinated expression between two photosynthetic petunia genes in transgenic plants. Mol. Gen. Genet. **211:** 507–514.

Göhring, F., and F. O. Fackelmayer, 1997 The scaffold/matrix attachment region binding protein hnRNP-U (SAF-A) is directly bound to chromosomal DNA *in vivo*: a chemical cross-linking study. Biochemistry **36:** 8276–8283.

Gygi, S. P., Y. Rochon, B. R. Franza and R. Aebersold, 1999 Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. **19:** 1720–1730.

Halpin, C., S. E. Cooke, A. Barakate, A. El Amrani and M. D. Ryan, 1999 Self-processing 2A-polyproteins—a system for coordinate expression of multiple proteins in transgenic plants. Plant J. **17:** 453–459.

Han, K. H., C. P. Ma and S. H. Strauss, 1997 Matrix attachment regions (MARs) enhance transformation frequency and transgene expression in poplar. Transg. Res. **6:** 415–420.

Holmes-Davis, R., and L. Comai, 1998 Nuclear matrix attachment regions and plant gene expression. Trends Plant Sci. **3:** 91–97.

Iida, S., O. M. Scheid, M. W. Saul, K. Seipel, C. Miyazaki *et al.*, 1992 Expression of a downstream gene from a bicistronic transcription unit in transgenic tobacco plants. Gene **119:** 199–205.

Jach, G., B. Görnhardt, J. Mundy, J. Logemann, E. Pinsdorf *et al.*, 1995 Enhanced quantitative resistance against fungal disease by combinatorial expression of different barley antifungal proteins in transgenic tobacco. Plant J. **8:** 97–109.

Jenuwein, T., W. C. Forrester, L. A. Fernandez-Herrero, G. Laible, M. Dull *et al.*, 1997 Extension of chromatin accessibility by nuclear matrix attachment regions. Nature **385:** 269–272.

Koncz, C., N. Martini, L. Szabados, M. Hrouda, A. Bachmair *et al.*, 1994 Specialized vectors for gene tagging and expression studies, pp. 1–22 in *Plant Molecular Biology Manual*, Vol. B2, Ed. 2, edited by S. B. Gelvin and R. A. Schilperoort. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Krohn, B. M., A. A. Hollier, S. Darchuk and D. M. Stark, 1998 Improving potato varieties through biotechnology. Asp. Appl. Biol. **52:** 239–244.

Laemmli, U. K., E. Käs, L. Poljak and Y. Adachi, 1992 Scaffold-associated regions: *cis*-acting determinants of chromatin structural loops and functional domains. Curr. Opin. Genet. Dev. **2:** 275–285.

Leech, M. J., K. May, D. Hallard, R. Verpoorte, V. de Luca *et al.*, 1998 Expression of two consecutive genes of a secondary metabolic pathway in transgenic tobacco: molecular diversity influences levels of expression and product accumulation. Plant Mol. Biol. **38:** 765–774.

Lough, T., C. Tourneur, J. Masson and Ch. Robaglia, 1997 Expression of genes in transgenic plants from bicistronic transcriptional units. Plant Sci. **129:** 91–99.

Ma, J. K. C., A. Hiatt, M. Hein, N. D. Vine, F. Wang *et al.*, 1995 Generation and assembly of secretory antibodies in plants. Science **268:** 716–719.

Maqbool, S. B., and P. Christou, 1999 Multiple traits of agronomic importance in transgenic indica rice plants: analysis of transgene integration patterns, expression levels and stability. Mol. Breed. **5:** 471–480.

Marcos, J. F., and R. N. Beachy, 1994 *In vitro* characterization of a cassette to accumulate multiple proteins through synthesis of a self-processing polypeptide. Plant Mol. Biol. **24:** 495–503.

Marcos, J. F., and R. N. Beachy, 1997 Transgenic accumulation of two plant virus coat proteins on a single self-processing polypeptide. J. Gen. Virol. **78:** 1771–1778.

Mlynárová, L., A. Loonen, J. Heldens, R. C. Jansen, P. Keizer *et al.*, 1994 Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. Plant Cell **6:** 417–426.

Mlynárová, L., R. C. Jansen, A. J. Conner, W. J. Stiekema and J.-P. Nap, 1995 The MAR-mediated reduction in position effect can be uncoupled from copy number-dependent expression in transgenic plants. Plant Cell **7:** 599–609.

Mlynárová, L., L. C. P. Keizer, W. J. Stiekema and J. P. Nap, 1996 Approaching the lower limits of transgene variability. Plant Cell **8:** 1589–1599.

Muskens, M. W. H., A. P. A. Vissers, J. N. M. Mol and J. M. Kooter, 2000 Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing. Plant Mol. Biol. **43:** 243–260.

Nap, J. P., W. G. Dirkse, J. Louwerse, J. Onstenk, R. Visser *et al.*, 1992 Analysis of the region in between two closely linked patatin genes: class II promoter activity in tuber, root and leaf. Plant Mol. Biol. **20:** 683–694.

Nap, J. P., M. van Spanje, W. G. Dirkse, G. Baarda, L. Mlynárová

et al., 1993a Activity of the promoter of the *Lhca3.St.1* gene, encoding the potato apoprotein 2 of the light-harvesting complex of photosystem I, in transgenic potato and tobacco plants. Plant Mol. Biol. **23:** 605–612.

Nap, J. P., P. Keizer and R. Jansen, 1993b First-generation transgenic plants and statistics. Plant Mol. Biol. Rep. **11:** 156–164.

Nap, J. P., A. J. Conner, L. Mlynárová, W. J. Stiekema and R. C. Jansen, 1997 Dissection of a synthesized quantitative trait to characterize transgene interactions. Genetics **147:** 315–320.

Nawrath, Ch., Y. Poirier and Ch. Somerville, 1994 Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in the high levels of polymer accumulation. Proc. Natl. Acad. Sci. USA **91:** 12760–12764.

Nayler, O., W. Strätling, J.-P. Bourquin, I. Stagljar, L. Lindemann *et al.*, 1998 SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. Nucleic Acids Res. **26:** 3542–3549.

Peach, C., and J. Velten**,** 1991 Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. Plant Mol. Biol. **17:** 49–60.

Phi-Van, L., and W. H. Strätling, 1988 The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. EMBO J. **7:** 655–664.

Poirier, Y., D. E. Dennis, K. Klomparens and Ch. Somerville, 1992 Polyhydroxybutyrate, a biodegradable thermoplastic, produced in transgenic plants. Science **256:** 520–523.

Poirier, Y., G. Ventre and C. Nawrath, 2000 High-frequency linkage of co-expressing T-DNA in transgenic *Arabidopsis thaliana* transformed by vacuum-infiltration of *Agrobacterium tumefaciens*. Theor. Appl. Genet. **100:** 487–493.

Poljak, L., C. Seum, T. Mattioni and U. K. Laemmli, 1994 SARs stimulate but do not confer position independent gene expression. Nucleic Acids Res. **22:** 4386–4394.

Razin, S. V., I. I. Gromova and O. V. Iarovaia, 1995 Specificity and functional significance of DNA interaction with the nuclear matrix: new approaches to clarify the old questions. Int. Rev. Cytol. **162B:** 405–448.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scheres, B., 2000 Non-linear signaling for pattern formation? Curr. Opin. Plant Biol. **3:** 412–417.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*, Ed. 3. W. H. Freeman, New York.

Strätling, W. H., and F. Yu, 1999 Origin and roles of nuclear matrix proteins. Specific functions of the MAR-binding protein MeCP2/ARBP. Crit. Rev. Euc. Gene Exp. **9:** 311–318.

Thompson, J. F., L. S. Hayes and D. B. Lloyd, 1991 Modulation of firefly luciferase stability and impact on studies of gene regulation. Gene **103:** 171–177.

Vain, P., B. Worland, A. Kohli, J. W. Snape, P. Christou *et al.*, 1999 Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J. **18:** 233–242.

van der Geest, A. H. M., G. E. Hall, Jr., S. Spiker and T. C. Hall, 1994 The phaseolin gene is flanked by matrix attachment regions. Plant J. **6:** 413–423.

van der Krol, A. R., and N.-H. Chua, 1991 The basic domain of the plant B-Zip proteins facilitates import of a reporter protein into plant nuclei. Plant Cell **3:** 667–675.

van Engelen, F. A., A. Schouten, J. W. Molthoff, J. Roosien, J. Salinas *et al.*, 1994 Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco. Plant Mol. Biol. **26:** 1701–1710.

van Engelen, F. A., J. W. Molthoff, A. J. Conner, J. P. Nap, A. Pereira *et al.*, 1995 pBINPLUS: an improved plant transformation vector based on pBIN19. Transg. Res. **4:** 288–290.

van Leeuwen, W., M. J. M. Hagendoorn, T. Ruttink, R. van Poecke, L. H. W. van der Plas *et al.*, 2000 The use of the luciferase reporter system for *in planta* gene expression studies. Plant Mol. Biol. Rep. **18:** 143–144 (full text available at http://www.nrc.ca/cisti/journals/ispmb/ispmb18/r00-022.pdf).

Vaucheret, H., T. Elmayan, D. Thierry, A. van der Geest, T. Hall *et al.*, 1998 Flank matrix attachment regions (MARs) from chicken, bean, yeast or tobacco do not prevent homology-dependent trans-silencing in transgenic tobacco plants. Mol. Gen. Genet. **259:** 388–392.

Ye, X., S. Al-Babili, A. Klöti, J. Zhang, P. Lucca *et al.*, 2000 Engineering the provitamin A (β-carotene) biosynthetic pathway into (caretonoid-free) rice endosperm. Science **287:** 303–305.

Zar, J. H., 1996 *Biostatistical Analysis*, Ed. 3. Prentice-Hall, New York.

Zhao, K., E. Kas, E. Gonzales and U. K. Laemmli, 1993 SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. EMBO J. **12:** 3237–3247.

Communicating editor: O. Savolainen